
Perfect Deterrence

Frank C. Zagare
and
D. Marc Kilgour



PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011-4211, USA www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain

© Frank C. Zagare and D. Marc Kilgour 2000

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2000

Printed in the United Kingdom at the University Press, Cambridge

Typeset in Palatino 10 / 12.5pt [c E]

A catalogue record for this book is available from the British Library

Library of Congress cataloguing in publication data

Zagare, Frank C.

Perfect deterrence / Frank C. Zagare, D. Marc Kilgour.

p. cm. – (Cambridge studies in international relations: 72)

Includes bibliographical references.

ISBN 0 521 78174 4 (hb) – ISBN 0 521 78713 0 (pb)

1. Deterrence (Strategy). I. Kilgour, D. Marc. II. Title. III Series.

U162.6.Z34 2000

355.02'17 – dc 21 99–088000

ISBN 0521 78174 4 hardback

ISBN 0521 78713 0 paperback

Contents

<i>List of figures</i>	<i>page</i> xiii
<i>List of tables</i>	xvi
<i>Preface</i>	xix
Part I Theoretical underpinnings	1
1 Classical deterrence theory	3
1.1 Classical deterrence theory: assumptions and implications	7
1.1.1 Structural deterrence theory	8
1.1.2 Decision-theoretic deterrence theory	16
1.2 Empirical anomalies	24
1.3 Logical problems	30
1.4 Coda	32
2 Rationality and deterrence	37
2.1 On rationality	38
2.1.1 Procedural rationality	38
2.1.2 Instrumental rationality	39
2.2 Some implications of the assumption of instrumental rationality	44
2.3 Resolving the paradox I: deterministic threats	49
2.4 Resolving the paradox II: the threat-that-leaves-something-to-chance	54
2.5 Resolving the paradox III: the metagame solution	58
2.6 Coda	63

Contents

3	Credibility and deterrence	65
3.1	On credibility	66
3.2	Resolving the paradox IV: mutually credible threats	70
3.3	On capability	81
3.4	Deterrence and uncertainty	84
3.4.1	Rudimentary Asymmetric Deterrence Game with complete information	86
3.4.2	Rudimentary Asymmetric Deterrence Game with one-sided incomplete information	87
3.5	Coda	93
	Part II Direct deterrence	97
4	Uncertainty and mutual deterrence	99
4.1	Capability, credibility, and mutual deterrence	101
4.2	Credibility, uncertainty, and mutual deterrence	104
4.3	Perfect Bayesian equilibria of the Generalized Mutual Deterrence Game	111
4.3.1	Class 1 equilibria	114
4.3.2	Class 2A and 2B Attack Equilibria	123
4.3.3	Class 3 Bluff Equilibrium	126
4.4	Coda	128
5	Unilateral deterrence	133
5.1	Game form	135
5.2	Unilateral deterrence under complete information	139
5.3	Unilateral deterrence under incomplete information	143
5.4	Perfect Bayesian equilibria of the Unilateral Deterrence Game	144
5.4.1	Deterrence equilibria	148
5.4.2	Other equilibria	153
5.5	The Unilateral Deterrence Game and power transition theory	158
5.6	Coda	164
	Part III Extended deterrence	167
6	Modeling extended deterrence	169
6.1	Preliminaries	170

6.2	The Asymmetric Escalation Game	175
6.3	Extended deterrence and the dynamics of escalation	183
6.4	The Asymmetric Escalation Game under complete information	187
6.4.1	Stability–instability paradox	189
6.4.2	Escalation dominance	192
6.5	Subgame-perfect equilibria	192
6.6	Coda	195
7	Modeling Massive Retaliation	197
7.1	Modeling all-or-nothing deployments	200
7.2	Perfect Bayesian equilibria under Massive Retaliation	203
7.2.1	Challenger-Soft Deterrence Equilibrium	205
7.2.2	No-Response Equilibrium	207
7.2.3	No-Limited-Response Equilibria	209
7.3	Coda	214
8	Modeling Flexible Response	219
8.1	Modeling Flexible Response: assumptions	221
8.2	Perfect Bayesian equilibria under Flexible Response	224
8.2.1	Escalatory Deterrence Equilibria	226
8.2.2	All-or-Nothing Equilibria	232
8.2.3	Flexible Response Equilibria	234
8.3	Multiple equilibria	243
8.4	Discussion	246
8.5	Coda	250
9	Limited war, crisis escalation, and extended deterrence	255
9.1	Theory and evidence	257
9.2	Assumptions	258
9.3	Behavioral possibilities	261
9.4	Deterrence and conflict spirals	262
9.4.1	Traditional deterrence	263
9.4.2	Non-traditional deterrence	266
9.4.3	Conflict spirals	270
9.5	Discussion	272
9.6	Coda	276

Contents

Part IV Implications	283
10 Perfect Deterrence Theory	285
10.1 Capability: a necessary condition	290
10.2 Deterrence and the status quo	293
10.3 Credibility and deterrence	296
10.4 Deterrence breakdowns, limited conflicts, and escalation spirals	301
10.5 Nuclear weapons and deterrence	303
10.6 Coda	307
Appendices	309
1 Deterrence models	311
2 Useful definitions	323
3 Rudimentary Asymmetric Deterrence Game	328
4 Generalized Mutual Deterrence Game	331
5 Unilateral Deterrence Game	339
6 Asymmetric Escalation Game: "Massive Retaliation" version	347
7 Asymmetric Escalation Game: "Flexible Response" version	354
8 Asymmetric Escalation Game: general version	365
<i>References</i>	384
<i>Index</i>	405

Figures

1.1	The weapons plane	<i>page</i>	10
1.2	Chicken as an informal model for deterrence		17
2.1	Extensive-form representation of Harsanyi's game		46
2.2	Strategic-form representation of Harsanyi's game		47
2.3	The game of Chicken		50
2.4	A first-level metagame of Chicken		59
2.5	A second-level metagame of Chicken (part)		61
3.1	Outcome matrix for the Generalized Mutual Deterrence Game		71
3.2	Extensive-form representation of the Generalized Mutual Deterrence Game		73
3.3	Ordinal strategic-form representation of Generalized Mutual Deterrence Game when neither player has a credible retaliatory threat		75
3.4	Ordinal strategic-form representation of Generalized Mutual Deterrence Game when both players have a credible retaliatory threat		78
3.5	Ordinal strategic-form representation of Generalized Mutual Deterrence Game when only one player (i.e., State A) has a credible retaliatory threat		80
3.6	Ordinal strategic-form representation of Generalized Mutual Deterrence Game when both players have a credible, but neither player has a capable, retaliatory threat		85
3.7	Rudimentary Asymmetric Deterrence Game		86
4.1	Generalized Mutual Deterrence Game		106
4.2	Outcome and utility notation for Generalized Mutual Deterrence Game		108

List of figures

4.3	Generalized Mutual Deterrence Game: location of equilibria	116
4.4	Generalized Mutual Deterrence Game: location of equilibria when the costs of unwanted conflict are high	125
5.1	Unilateral Deterrence Game	136
5.2	Outcome and utility notation for Unilateral Deterrence Game	137
5.3	Determining subgame-perfect equilibrium in the Unilateral Deterrence Game with credible threats and complete information using backward induction	141
5.4	Unilateral Deterrence Game with incomplete information: location of equilibria	145
6.1	Containment <i>circa</i> 1947	174
6.2	Massive Retaliation <i>circa</i> 1952	179
6.3	Asymmetric Escalation Game	181
6.4	Strategic-form representation of Asymmetric Escalation Game	184
6.5	Outcome array of Asymmetric Escalation Game	185
6.6	Analysis of Asymmetric Escalation Game when threats are incredible	188
6.7	Outcome array of Asymmetric Escalation Game when threats are incredible	189
6.8	Preferences associated with the stability–instability paradox	191
6.9	Preference assumptions illustrating escalation dominance	193
7.1	Asymmetric Escalation Game when Defender lacks a credible first-level threat	204
7.2	Location of No-Response Equilibrium and No-Limited-Response Equilibria	208
8.1	Extended-deterrence deployment strategies within the strategic concept of Flexible Response	221
8.2	Asymmetric Escalation Game under Flexible Response	222
8.3	Existence conditions for Limited-Response Deterrence Equilibria	236
8.4	Existence conditions for No-First-Use and Warfighting Escalatory Limited-Response Equilibria	241
9.1	Existence regions for equilibria of the Spiral Family	274

List of figures

A1.1	Rudimentary Asymmetric Deterrence Game	313
A1.2	Utilities in the Rudimentary Asymmetric Deterrence Game	314
A1.3	Generalized Mutual Deterrence Game	315
A1.4	Utilities in the Generalized Mutual Deterrence Game	316
A1.5	Unilateral Deterrence Game	318
A1.6	Utilities in the Unilateral Deterrence Game	319
A1.7	Asymmetric Escalation Game	320
A1.8	Utilities in the Asymmetric Escalation Game	321
A3.1	Rudimentary Asymmetric Deterrence Game	329
A4.1	Generalized Mutual Deterrence Game	332
A5.1	Unilateral Deterrence Game	340
A6.1	Asymmetric Escalation Game: “Massive Retaliation” version	348
A7.1	Asymmetric Escalation Game: “Flexible Response” version	355
A8.1	Asymmetric Escalation Game: general version	367

Tables

3.1	Forms of subgame-perfect equilibria for the Rudimentary Asymmetric Deterrence Game with complete information	<i>page</i> 88
3.2	Forms of equilibria and existence conditions for the Rudimentary Asymmetric Deterrence Game with one-sided incomplete information	91
4.1	Mutual deterrence under complete information	102
4.2	Perfect Bayesian equilibria and existence conditions for Generalized Mutual Deterrence Game with incomplete information	114
5.1	Subgame-perfect equilibria for Unilateral Deterrence Games with complete information	140
5.2	Perfect Bayesian equilibria and existence conditions for the Unilateral Deterrence Game with incomplete information	147
6.1	Cases of immediate extended deterrence among major powers, 1895–1985	172
6.2	Subgame-perfect equilibrium outcomes in Asymmetric Escalation Game	194
7.1	Perfect Bayesian equilibria and existence conditions for Asymmetric Escalation Game when Defender’s first-level threat lacks credibility	205
8.1	Perfect Bayesian equilibria of the Asymmetric Escalation Game when Defender’s first-level threat is completely credible	227
8.2	Perfect Bayesian equilibria and associated deployment policies	228

List of tables

8.3	Summary of preferences when equilibria overlap in Asymmetric Escalation Game (Flexible Response version)	244
9.1	Types and credibility parameters	260
9.2	Equilibria of the Asymmetric Escalation Game when Challenger has high credibility	264
10.1	Crisis structures and outcomes	299
A1.1	Four deterrence models	312
A4.1	Subgame-perfect Bayesian equilibria of the Generalized Mutual Deterrence Game	334
A5.1	Perfect Bayesian equilibria of Unilateral Deterrence Game with incomplete information	345
A8.1	Perfect Bayesian equilibria of Asymmetric Escalation Game with incomplete information	370

1 Classical deterrence theory

International relations is a study that is plagued with platitudes.

A.F.K. Organski

For over forty-five years, the rivalry between the United States and the Soviet Union defined the world we live in. Civil wars in Africa, *coups d'état* in Latin America, revolutions in Asia, and small wars around the globe were filtered through the prism of the Cold War, not only in Washington and Moscow, but in just about every major capital on the planet.¹

The global contest between the superpowers was both dramatic and dangerous. As is generally the case in hegemonic competitions, the stakes were high: control of the international system lay in the "balance." But for some, and later most, strategic thinkers, the dropping of atomic bombs on Hiroshima and Nagasaki instantly and permanently changed the nature of the international system and the laws that govern it. Bernard Brodie was the first to argue that the world before 1945 was fundamentally different from the world that would follow. Up to that point, he argued, "the chief purpose of our military establishment [had] been to win wars. From now on its chief purpose must be to avert them. It can have almost no other purpose" (Brodie, 1946: 76).

If the post-World War II world were truly *sui generis*, as Brodie and others argued, then a new theory would be needed to replace the conventional wisdom of the past. The enormous costs associated with warfare after 1945 would clearly be the cornerstone of this new theory. But there was another essential difference between the older and the

¹ This chapter is based on Zagare (1996a).

Theoretical underpinnings

newer world order that any new theory would have to take account of.

Prior to 1939, the international system was decidedly multipolar as several great states, and a handful of lesser states, vied for power and influence around the globe. But after the defeat of Germany and Japan in 1945, this was no longer the case. The multipolar Eurocentric world had suddenly been transformed into a system dominated by two superpowers from the periphery of the European state system. The bipolar nature of the post-war period would also have to be considered by theorists trying to understand the inner workings of the new system.

It was in this context that *classical* (or *rational*) *deterrence theory* was born. Although Brodie is considered its father, the theory had a long and distinguished pedigree; as we will show, the conceptual break with past intellectual traditions was not as complete as is sometimes claimed.

As classical deterrence theory matured in the 1950s and early 1960s, many strategic thinkers nurtured its growth. Scholars like Herman Kahn, Thomas Schelling, Albert Wohlstetter, Oskar Morgenstern, William Kaufmann, and Glenn Snyder contributed mightily to its development and refinement. In time, the theoretical edifice they created came to be seen as the Rosetta Stone of the nuclear age. As a descriptive tool, it was used to explain the operation of the international system and its constituent parts; and, as a normative device, policy-makers in the United States and later the Soviet Union employed it as a guide to action. With seemingly good reason, the tenets of the theory became, in both academic and official circles, the conventional wisdom. Not only did classical deterrence theory purport to explain the absence of a US-USSR war after 1945 but, if properly heeded, could be used to all but eliminate the possibility of future superpower conflicts.

Deadly nuclear weapons and a carefully maintained strategic balance were the “twin pillars” upon which this global nirvana rested (Gaddis, 1986; Waltz, 1993). Each was seen as a necessary condition for peace and stability.² Thus, the superpowers were simultaneously

² As Levy (1985: 44) rightly observes, “‘stability’ is one of the more ambiguous concepts in the international relations literature.” At one time Waltz (1964) equated stability with peace, and instability with war. But his definition left open the critical question of how to treat periods of crisis. As Lebow (1981) notes, crises fall between peace and war. This is perhaps why Mearsheimer (1990: 7) defines stability “as the absence of

enjoined not to “build down” by dramatically reducing their ability to inflict unacceptable damage on one another, and not to “build up” by seeking unilateral advantage. If ever there was a theory that enshrined the status quo, this was it.

In 1989 the Berlin Wall was torn down. A few months later the Soviet empire in Eastern Europe dissolved. By 1991 the Soviet Union itself was in tatters. Amidst the euphoria and astonishment that surrounded these events, many observers were sure that this *new* new world order, even as it evolved, would be inordinately peaceful. Indeed, some were ready to proclaim the “end of history” and, with it, the eradication of ideological struggles (Fukuyama, 1992). What was overlooked by all but a few (e.g., Huntington, 1989; Mearsheimer, 1990), however, was that this putatively ultra-stable environment differed markedly from the bipolar nuclear system that had been so widely credited with maintaining peace since the end of World War II.

Was the Cold War period, particularly after the Soviet Union achieved nuclear parity with the United States, as stable as classical deterrence theorists claimed? If so, it follows that the collapse of the Soviet Union was a destabilizing event that will shortly make us long for the “good old days” of the Cold War (Mearsheimer, 1990). But if not, then another new theoretical framework is needed, not only to explain the relative tranquillity of the past, but also to understand and manage the present and future world. In particular, if the bipolar nuclear relationship of the superpowers was as dangerous as some now assert, then the actual stability of the Cold War era remains the anomaly to be explained.

To cut to the chase: this book argues that classical deterrence theory is flawed, both empirically and logically. Moreover, this book seeks to provide a theoretical framework – Perfect Deterrence Theory – from which to view the world we are living in now. As well, by offering an explanation of the workings of bilateral conflict relationships, it attempts to come to grips with the old world order so recently left behind.

wars *and* major crises” (emphasis added). Later, Waltz (1993: 45) redefined the concept in terms of systemic durability: “systems that survive major wars thereby demonstrate their stability.” To eliminate possible confusion we shall follow Mearsheimer and restrict our use of the term as follows: when we say that either a system or a deterrence relationship is stable, we mean that the status quo is likely to survive; and when we say that a system or a deterrence relationship is unstable, we mean to imply that *either* a crisis or a war is possible.

Theoretical underpinnings

Unlike classical deterrence theory, Perfect Deterrence Theory is not confined to relationships between nuclear states. It is perhaps understandable that each generation of statesmen (and academics) sees its own era as unique. But this particular conceit, however comforting to those wishing to avoid the mistakes of their forebears, does not stand up to unimpassioned scrutiny. There is simply no compelling reason to believe that the prohibitively high costs of conflict are likely to inoculate contemporary states against warfare, no more so than there was when Sir Norman Angell (1910) made this very same argument just prior to World War I.

Because Perfect Deterrence Theory affords no special status to nuclear weapons, its logical and empirical domain is not confined to superpower relationships. Of course, this is not to say that nuclear weapons are necessarily “irrelevant,” as Mueller (1988) and a few others have suggested. Rather our position is that if weapons – nuclear or otherwise – that alter the costs of war have an impact, we hope to ascertain what that impact is, and when it comes into play. But we do not believe that particularly powerful weapons necessarily require a theory unto themselves.

Put in a slightly different way, Perfect Deterrence Theory is completely general and should apply as well to conflict-of-interest situations between various combinations of large and small states, with or without nuclear capabilities. In fact, with the proper modifications and provisos, the set of interrelated models we develop may be used to explore contentious relationships between non-state actors, between organized groups, or even between individuals. We hold that the underlying dynamic of human strife, however aggregated, is fundamentally the same.

Nonetheless, because interstate conflict remains our principal focus, classical deterrence theory will be our point of departure. As Kenny (1985: ix) notes, “deterrence is the key concept for the understanding of the strategy and diplomacy of the age.” And, as DeNardo (1995: 2) astutely observes, “as long as weapons of mass destruction and hostile relationships coexist in world politics, the question of deterrence will not go away.” We begin, therefore, with a description of the underlying premises and principal conclusions of classical deterrence theory, detailing along the way some of its logical and empirical deficiencies.

1.1 Classical deterrence theory: assumptions and implications

Because there is no single, authoritative exposition of its major premises, an outline of classical deterrence theory must be pieced together from a variety of sources. Fortunately, there is wide consensus among theorists on both the provenance and the broad contours of the theory.³ It is generally agreed that the roots of classical deterrence lie in the intellectual tradition that has variously been labeled “political realism,” “*realpolitik*,” or “power politics.” This state-centric approach – which some trace back to Thucydides or earlier – posits egoistic, rational, and undifferentiated⁴ units driven by their nature to maximize power (Morgenthau, 1948), or by their environment to maximize security (Waltz, 1979). When aggregated, these units constitute a self-help system that resembles Hobbes’s “state of nature” where the life of man is “solitary, poore, nasty, brutish, and short.” In the realists’ paradigm, the international system, like Hobbes’ anarchistic pre-societal state, is seen to lack an overarching authority or sovereign (Milner, 1991). Thus, each state in the system must “rely on [its] own strength and art for caution against all others” (Hobbes, 1968 [1651]: 224).

In a system where every state must provide for its own security, most realists hold that a *balance of power* is the most efficient mechanism for maintaining order (Morgenthau, 1948; Claude, 1962; Waltz, 1993; Kissinger, 1994). When power is equally distributed among actors in the system, or among the major partitions of actors – as the argument goes – peace is more likely since no one state has an incentive to upset the status quo and challenge another. By contrast, an asymmetric distribution of power provides no check on stronger states intent on enhancing their welfare. Or as Mearsheimer (1990: 18) puts it, “power inequalities invite war by increasing the potential for successful aggression; hence war is minimized when inequalities are least.”

³ Glaser (1989) subdivides the wider strategic literature into three categories. What we call classical deterrence theory corresponds most closely to what Glaser terms the punitive retaliation school.

⁴ This is one reason why Waltz (1993: 47) goes out of his way to argue that “our conviction that the United States was the status quo and the Soviet Union the interventionist power distorted our view of reality.” For Waltz and other classical deterrence theorists, *all* states are essentially the same; they are insecure, afraid, and protective of their vital interests.

Theoretical underpinnings

Classical deterrence theory builds upon this theoretical base, and extends its domain, by considering the consequences of war in the nuclear age. In this regard, two distinct, yet compatible, strands of the theory can be discerned: *structural* (or *neorealist*) *deterrence theory* (Kaplan, 1957; Waltz, 1979; Mearsheimer, 1990) and what we shall refer to, for want of a better term, as *decision-theoretic deterrence theory* (Ellsberg, 1959, 1961; Schelling, 1960, 1966; Jervis, 1972; Snyder, 1972). As Allison (1971) clearly demonstrates, both of these complementary approaches to deterrence share a conceptual orientation with the realist approach to international politics. In the strategic literature, these two strands converge to form the pastiche of classical deterrence theory.

1.1.1 Structural deterrence theory

Like traditional balance of power theorists, structural deterrence theorists see the key to international stability in the distribution of power – within the system, in general, and among the great powers in particular. Most structuralists hold that when a parity relationship is combined with the enormous absolute costs of nuclear war, a deliberate (i.e., a “rational”) war is at once unthinkable and virtually impossible. Those who subscribe to this view see the nuclear balance as unusually robust and stable, and credit the absence of a major superpower conflict in the post-war period directly to the enormous destructive power of nuclear weapons.

Structural deterrence theorists offer numerous explanations for the pacifying impact of bipolar structures. Waltz (1964: 882–886), for instance, argues that “the remarkable stability” of the bipolar post-war period derives from the absence of peripheries, the intensity of the competition, the “nearly constant presence of pressure and the recurrence of crisis,” and the preponderant power of the two major contenders. Gaddis (1986: 105–110) cites different factors. For Gaddis, the “long-peace” of the post-war period can be traced to a “relatively simple structure” that at once reflected the realities of power, was easy to maintain, had a more stable alliance system, and could easily absorb shifts in alliance patterns. Mearsheimer (1990: 14) sees still other reasons why bipolar systems are, in general, more peaceful than multipolar systems: “First, the number of conflict dyads is fewer, leaving fewer possibilities for war. Second, deterrence is easier, because imbalances of power are fewer and more easily averted. Third, the prospects for deterrence are greater because miscalcula-

tions of relative power and of opponents' resolve are fewer and less likely."

While structural deterrence theorists may differ about exactly why bipolar systems are inherently more peaceful than multipolar systems, they are in almost unanimous agreement about the consequences of nuclear weapons.⁵ Virtually every structural deterrence theorist believes that the high cost of war in the nuclear era has rendered states more prudent and, simultaneously, raised the provocation level necessary for outright conflict (Snyder and Diesing, 1977: 450–453). When these effects are combined with the pacifying tendencies of a bipolar system, a world order is produced that, *when properly managed*, is unlikely to be characterized by major interstate war.

Of course, this judgment is subject to many qualifications and provisos. Most important is the nature of current military technology. Specifically, when defensive aspects are ascendant, or are thought to be ascendant, the underlying stability of a parity relationship, be it bipolar or multipolar, is reinforced. But when offensive aspects predominate, as was *believed* to be the case in 1914, even a strict bipolar structure could witness war (Wohlstetter, 1959; Quester, 1977; Jervis, 1978; Snyder, 1984; Van Evera, 1984: 72).

The intricate relationship between system structure, the cost of war, and the characteristics of weapons systems is succinctly captured in a formal model of a missile war developed by Intriligator and Brito (1984, 1987). Since this model reflects the underlying assumptions of structural deterrence theory and highlights several of its important and non-obvious implications, we shall use its original graphical representation (figure 1.1) as an organizing device for summarizing the principal tenets of structural deterrence theory.⁶ We realize that our tack is fraught with danger: we risk oversimplifying an extensive literature characterized by nuance and subtlety.⁷ Nonetheless, undeterred, we shall proceed according to this plan. The reader should

⁵ A concise summary of the debate about the war proneness of different systemic configurations can be found in Kegley and Raymond (1994). See also Sabrosky (1985).

⁶ For an incomplete information game model that reaches similar conclusions about the conditions of war and peace, see Bueno de Mesquita, Morrow, and Zorick (1997). In this model, which is based on assumptions that are compatible with classical deterrence theory, the probability that the status quo will be challenged increases as the observable military advantage of one side or the other increases.

⁷ See, for instance, the wide range of responses (and non-responses) to Vasquez's penetrating evaluation of the realist paradigm in the December 1997 issue of the *American Political Science Review*.

Theoretical underpinnings

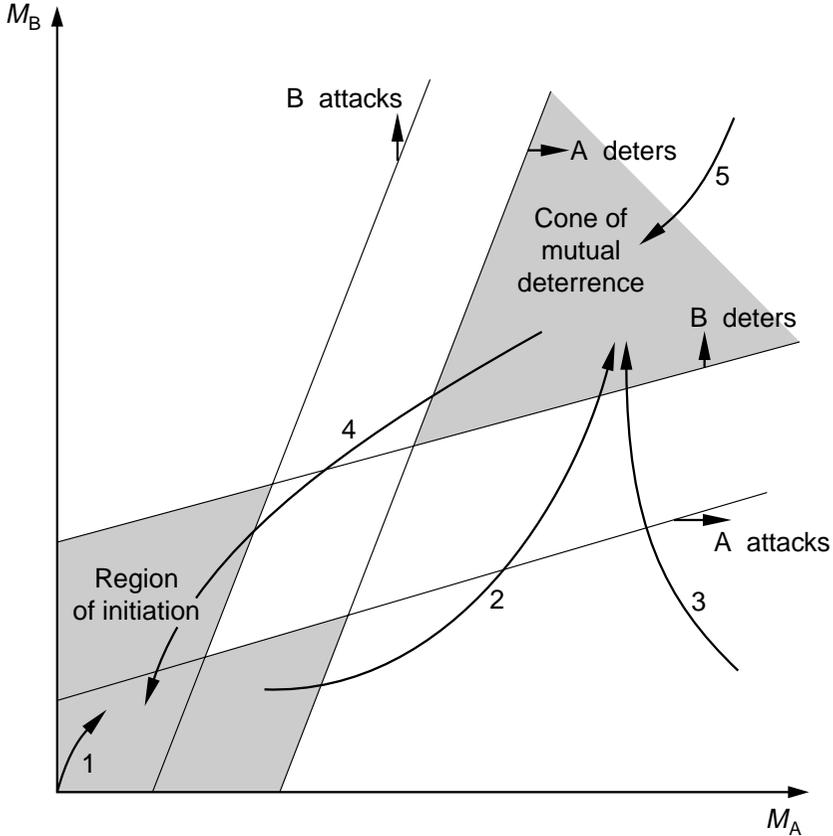


Fig. 1.1. The weapons plane. (Source: Intriligator and Brito, 1984: 77.)

keep in mind, though, that we are attempting to offer a *consensus* view of a diverse and multifaceted approach to international conflict.⁸

As one might expect, there are two actors (here called State A and State B) in this model. Like the individuals in Hobbes' state of nature, the states are linked in a hostile relationship characterized by mutual mistrust and fear. Both states are rational and their relationship is governed by the absolute cost each is capable of imposing on the other in a conflict. Note that the states are undifferentiated units: each considers attacking the other. They are also egotistical: each takes into

⁸ For a balanced discussion of the varieties of structural realist thought, see Elman (1996).

account the costs it can impose on the other only to the extent that those costs alter the other's behavior and, consequently, its own payoff. Ethical, moral, and legal considerations do not enter into either state's decision-making framework.

War costs are determined by the number and characteristics of the weapons in each state's arsenal, and by the strategic choices of each state's decision-makers. Strategic decisions involve two critical choices: "First there is a choice of targets between counterforce targets of enemy weapons and countervalue targets of enemy cities and industrial capability . . . Second, there is a choice of rate of fire between the extreme values of a maximum rate, i.e., firing all weapons as rapidly as possible, and a zero rate, i.e., holding weapons in reserve for later use" (Intriligator and Brito, 1987: 15). A state's choice of a target and rate of fire together constitute its *grand strategy*.⁹

In the model, the two states choose optimal grand strategies that will inflict certain costs on one another when and if a war occurs. The anticipation of these costs, in turn, drives each state's behavior. In particular, State A *attacks* State B "if it can launch a first strike on B . . . and reduce the number of B weapons sufficiently that B does not have enough weapons left to inflict unacceptable casualties on A in a massive retaliation strike." By contrast, State A *deters* State B "if given a first strike by B . . . A can absorb this strike and have enough weapons left to inflict an unacceptable level of casualties to B in a retaliatory massive retaliation strike" (Intriligator and Brito, 1987: 16, 18). Notice that the incentive to attack is presumed. Consequently, there is only one determining variable in each state's calculus – the cost of attacking.

Given optimal targeting and firing strategies, the costs and benefits of attacking depend on the absolute number of weapons in both arsenals. The consequences of the states' cost/benefit calculations can be determined by locating their combined arsenals in the *weapons plane* of figure 1.1. In this figure, the number of missiles available to State A, M_A , is measured along the horizontal axis, while the vertical axis measures the number of missiles possessed by State B, M_B .

Notice the four critical thresholds represented by the two sets of parallel lines. One line in each set represents the cost level beyond which A or B is deterred; the other represents the point below which

⁹ Wagner (1991) uses game theory to evaluate the strategic implications of counterforce targeting options.

Theoretical underpinnings

benefits outweigh costs so that either A or B is motivated to attack. The intersections of these lines define nine different regions in the weapons plane. (Ignore for now the curved arrows indicating movements on the plane.) The behavioral patterns anticipated in each region constitute the principal conclusions of the model. These conclusions are congruent with the major tenets of structural deterrence theory.¹⁰

1. *Parity relationships, when coupled with high war costs, are especially peaceful.* When war costs are mutually high, bilateral strategic relationships fall into the shaded region in the northeast section of figure 1.1 – called the *cone of mutual deterrence*. Clearly, relations within this cone reflect the MAD (Mutual Assured Destruction) condition: each state can inflict unacceptable costs on the other, regardless of which attacks first. All structural deterrence theorists (by definition), and most balance of power theorists, hold that as soon as this condition is met, peace is at hand (see, *inter alia*, Kaufmann, 1956; Brodie, 1959; Snyder, 1961; Glaser, 1990; Mearsheimer, 1990).

By contrast, when the cost of outright conflict is low, even parity may be insufficient to preclude confrontations, suggesting that “war is always possible among states armed only with conventional weapons” (Waltz, 1993: 77). In figure 1.1, the sawtooth-shaped region adjacent to the origin represents all strategic relationships characterized by low war costs. For obvious reasons, Intriligator and Brito call it the *region of initiation*. In the center section of the region, where parity reigns, neither side can deter the other; conflict is almost inevitable. Thus, “this portion is one of virtually forced preemption in which it is greatly advantageous to initiate rather than retaliate. The ‘reciprocal fear of surprise attack’ based on the tremendous advantage in striking first forces both sides to initiate, each trying to preempt the attack of the other” (Intriligator and Brito, 1984: 73–74).

2. *Asymmetric power relationships are associated with crises and war.* The most precarious form of asymmetry occurs when neither state can deter the other (i.e., when war costs are mutually low), but when one of them nonetheless calculates an advantage in attacking first. Thus, along with the center portion of the region of initiation (see above), both the lower right and upper left portions exhibit “instability against war outbreak” (Intriligator and Brito, 1984: 74).

Still, even when one state can deter the other, some form of conflict

¹⁰ For the sensitivity of these results to the model’s initial conditions, see Mayer (1986).

is likely. This conclusion follows from the assumption that all states have similar concerns and motivations (i.e., states are undifferentiated units). Consequently, general deterrence is unlikely to succeed¹¹ in the entire area of the weapons plane below the line labeled "A attacks" and to the left of the line labeled "B attacks." Whenever at least one state is undeterred, war remains a distinct possibility. Of course, outright conflict *might* be averted if the disadvantaged state follows a policy of appeasement and adjusts its policies to reflect the stronger state's interests (Kugler and Zagare, 1990: 60–63). In either case, though, the status quo is unlikely to survive.

3. *As the absolute costs of war increase, ceteris paribus, the probability of war decreases.* Or in Mearsheimer's (1990: 19) words, "the more horrible the prospect of war, the less likely it is to occur." In fact, with an overkill capability that places both states deep within the cone of mutual deterrence, the probability of war "may be reduced to virtually zero" (Intriligator and Brito, 1981: 256).

The functional relationship between war costs and war proneness can be observed by considering the strategic implications of various trajectories (represented by curved arrows) through the weapons plane of figure 1.1. As one moves northeast away from the region of initiation (trajectory 2) where each state can attack the other, or away from either asymmetric area in which only one state can attack (for

¹¹ Like "stability" (see footnote 2), deterrence "success" is an "essentially contested" concept (MacIntyre, 1973). One reason is that success and failure are relative terms (Levy, 1988: 498). If deterrence success is equated with the absence of war, deterrence can be said to have succeeded even when a crisis occurs, or when one state is able to win concessions from another by threatening war. Huth's (1988a: 25) coding scheme is consistent with this conceptualization. For example, Huth codes the Berlin crisis of 1948 and the 1954–55 confrontation between China and the United States over Quemoy and Matsu as successes. Since the domain of Huth's empirical study is restricted, in part, to "immediate" deterrence encounters in which one state is "seriously considering attacking" another (Morgan, 1977: 33), his definition is understandable. But in another sense, deterrence failed: a challenge occurred. To account for these subtleties, we reserve the term "deterrence success" to indicate situations in which the status quo is not disturbed. We use the terms "deterrence success" and "general deterrence success" synonymously. (The term "general deterrence" refers to hostile relationships in which no state "is anywhere near mounting an attack" against its opponent [Morgan, 1977: 28].) We take "immediate deterrence success" to imply that the status quo has been contested but that an all-out conflict (e.g., war) has been avoided. Clearly, immediate deterrence cannot succeed unless general deterrence has failed. Finally, we equate an "immediate deterrence failure" with all-out conflict. In chapter 9, we relax our qualifications to take into account a wider range of outcomes than we currently consider.

Theoretical underpinnings

instance, trajectory 3), toward the cone of mutual deterrence where war costs are mutually prohibitive, the probability of war initiation decreases dramatically.¹² By contrast, movement downward through the cone (see trajectories 4 and 5) toward the origin of the weapons plane and the region of initiation (where war costs are lower) only increases the prospect of conflict, especially when such bilateral disarmament is “carried too far” (Intriligator and Brito, 1987: 22).

In sum, structural deterrence theory attributes the “long-peace” of the Cold War era to the *balance of terror*, that grizzly combination of rough parity and high destructiveness unique to the nuclear age. From this axiom flow several practical, policy-orientated, conclusions.

- *First, quantitative arms races, which serve to increase the cost of conflict, can help prevent wars* (Gray, 1974: 209).

Like the Roman military strategist Vegetius, then, most classical deterrence theorists hold that proper preparation for war reduces its likelihood. For this reason, they worked against the nuclear freeze movement during the 1980s, opposed all bilateral – or worse – unilateral efforts to disarm during the Cold War and, in the early days of the Reagan era, favored a shift to single-warhead missiles (Art, 1985; Aspin, 1986).

- *By contrast, qualitative arms races, which threaten to provide one side or another with a first-strike advantage, increase the probability of preemptive war.*¹³

This is one reason why most classical deterrence theorists opposed the development of more accurate delivery systems, resisted the deployment of missiles armed with multiple warheads (Jervis, 1978; Scoville, 1981), and disputed the implementation of counterforce targeting doctrines (Van Evera, 1984). Also related were worries about

¹² Trajectory 1, which begins at the origin of figure 1.1, is an exception to this statement. Clearly, when neither state has any weapons, outright conflict is impossible. Thus, any movement away from the origin into the region of initiation raises the probability of war.

¹³ For a contrary view, see Huntington (1958). It is worth mentioning, however, that Huntington’s argument that quantitative arms races increase the probability of war rests upon the supposition that they lead to an asymmetric distribution of power. In Huntington’s view, qualitative arms races reinforce parity and, therefore, reduce war’s likelihood.

a “window of vulnerability” in the late 1970s and early 1980s (Nitze, 1976/77; Johnson, 1983).

- *Comprehensive and effective defense systems make conflict more likely.*

Because of the purported consequences of strategic defense systems, most classical deterrence theorists in the United States worked against the Strategic Defense Initiative (or “Star Wars” program) of the Reagan administration (Bundy *et al.*, 1984/85), continue to oppose anti-ballistic missile (ABM) systems, and support the SALT I Treaty that first limited and eventually abolished them (National Academy of Sciences, 1997: 42–46).

- *The selective proliferation of nuclear weapons can help prevent war and promote peace.*

Although often unappreciated, this argument rests on, and indeed flows from, the supposition that nuclear deterrence is “very robust” (Berkowitz, 1985: 117). Recognizing the logical connection between premise and conclusion, many classical deterrence theorists have supported the “prudent” dissemination of nuclear technology.¹⁴ Early in the nuclear age, for instance, Oskar Morgenstern (1959, 74–77) submitted that it was in the interest of the United States to provide the Soviet Union with an invulnerable strategic retaliatory force,¹⁵ and Pierre Gallois (1961) defended the French decision to go nuclear because he believed that an independent nuclear force would reduce French vulnerability to political coercion and outright attack.

More recently, Mearsheimer (1990: 54) has suggested that a German nuclear capability is “the best hope for avoiding war in post-Cold War Europe” and argued that peace is much more likely if Ukraine retains its nuclear force (Mearsheimer, 1993). And, if former Soviet spymaster Pavel Sudoplatov (1994: 195) is to be believed, it was precisely the logic of proliferation that led several of the West’s leading scientists to pass

¹⁴ While this is a minority position, it is telling that support for controlled proliferation policies comes from writers (e.g., Waltz and Mearsheimer) whose work is most explicitly theoretical.

¹⁵ Morgenstern was particularly concerned with the impact of an asymmetric strategic relationship that, paradoxically, might cause a weaker, more vulnerable state (i.e., the Soviet Union) to preempt a stronger, less vulnerable state (i.e., the United States). Note, however, that Morgenstern presumed that the chance of war between two states possessing an invulnerable second-strike nuclear capability was remote.

Theoretical underpinnings

sensitive information derived from the Manhattan Project to the Soviets: “Since [J. Robert] Oppenheimer, [Niels] Bohr and [Enrico] Fermi were fierce opponents of violence, they would seek to prevent a nuclear war, creating a balance of power through sharing the secrets of atomic energy.” Waltz (1981), Intriligator and Brito (1981), Bueno de Mesquita and Riker (1982), Van Evera (1990/91), and Posen (1993), *inter alia*, have also made the case for the discriminate distribution of nuclear technology.¹⁶

- *Accidental war is the gravest threat to peace.*

Here, too, the argument rests on the theoretical consequences of the confluence of parity and the high costs of nuclear war (see, for example, Morgenstern, 1959: 69; Smoker and Bradley, 1988). When both conditions are present, a deliberate (i.e., rational) war is improbable; hence, nuclear wars are most likely to occur by mistake (Abrams, 1988; Intriligator and Brito, 1981; Sagan, 1993; Brito and Intriligator, 1996). To prevent accidental nuclear war, classical deterrence theorists argue for redundancy in command, control, communication, and intelligence (C³I) systems (Bracken, 1983) and against “launch on warning” doctrines (Blair, 1993: 174).

1.1.2 Decision-theoretic deterrence theory

Unlike structural deterrence theory, which finds the key to interstate stability in the structure and distribution of power, decision-theoretic deterrence theory focuses on the interplay of outcomes, preferences, and choices in determining interstate conflict behavior. The genre includes both formal and informal rational choice (expected utility) analyses and subsequent game-theoretic refinements. In the discussion that follows, we lump all these methodologies together, ignoring important differences. Right now our purpose is simply to highlight their common theoretical point of view.¹⁷

¹⁶ Structural deterrence theorists do not, however, favor disseminating nuclear weapons to “crazy states” or their “irrational leaders.” As we note below, however, this escape clause is inconsistent with two fundamental axioms of their approach; namely, the assumption that states are undifferentiated and rational. Thus, like Great Britain during the eighteenth and nineteenth centuries (Organski, 1958), Iraq, Libya, Iran, Cuba, and North Korea appear to be the current exceptions to the rule that *all* states are self-interested power maximizers.

¹⁷ Our label for this group of theorists could be misleading. We do not include all expected utility and game-theoretic models of deterrence in this category, only those that share the modal assumptions discussed below.

		State B	
		Cooperate (C)	Defect (D)
State A	Cooperate (C)	<i>Status Quo</i> (3,3)	<i>B Wins</i> (2,4)
	Defect (D)	<i>A Wins</i> (4,2)	<i>Conflict</i> (1,1)

(x,y) = payoff to A, payoff to B
 4 = best; 3 = next-best; 2 = next-worst; 1 = worst

Fig. 1.2. Chicken as an informal model for deterrence.

Beginning where structural theorists leave off, the decision-making strand of classical deterrence theory posits a situation in which nuclear war is so costly that only an “irrational” leader could consider it a means of conflict resolution. Thus, a critical *deduction* of structural deterrence theory is accepted and embedded as an *axiom* by decision-theoretic deterrence theorists.

Since nuclear war was taken to be at once irrational and unthinkable, interstate crises came to be seen as its functional equivalent, that is, a contest for exhibiting and measuring power (Waltz, 1964: 884; Hoffman, 1965). To represent war’s surrogate, most classical deterrence theorists used, as a formal or informal metaphor, the deceptively simple game of *Chicken* (or a structural equivalent).¹⁸

Chicken is a stark model of the interaction of two decision-makers.

¹⁸ As is well known, this game models a contest reportedly indulged in by reckless teenagers who would drive cars toward each other at high speed. The first driver to swerve was the “chicken,” and was disgraced. Of course, not swerving was much worse – for both drivers. (For applications of Chicken to deterrence, see, *inter alia*, Kahn, 1960, 1962, 1965; Snyder, 1971; Hopkins and Mansbach, 1973; Brams, 1975, 1985; Jervis, 1979; Powell, 1987, 1990; Nicholson, 1989; and especially, Schelling, 1960, 1966.)

Theoretical underpinnings

As before, call them State A and State B. Each state is seen as having two broad strategic choices: either to *cooperate* (C) by supporting the status quo, or to *defect* (D) from cooperation by seeking to overturn it. These alternative choices (or strategies) give rise to four broad outcomes: if both states cooperate, the *Status Quo* (outcome SQ) prevails; if one state cooperates and the other does not, the non-cooperator wins or gains an advantage (either *A Wins* [outcome DC] or *B Wins* [outcome CD]); and if neither state cooperates, *Conflict* (for now, read “nuclear war”) occurs (outcome DD). By definition, in Chicken, each player most prefers to gain the advantage and “win” the game, next prefers mutual cooperation (i.e., the *Status Quo*), next prefers to concede the advantage to its opponent and, significantly, least prefers *Conflict*.¹⁹

The strategies, outcomes, and ordinal rankings are summarized in figure 1.2, which for now we treat as an informal model rather than as a *strategic-form* game.²⁰ Preference rankings are represented by an ordered pair in each cell of the matrix that indicates State A’s (row’s) and State B’s (column’s) preference ranking of the four outcomes. The most-preferred outcome is indicated by a rank of 4, the next most-preferred by 3, and so on. For example, in Chicken, outcome DC (*A Wins*) is State A’s best outcome (i.e., rank 4 for A) and State B’s next-worst outcome (i.e., rank 2 for B).

Chicken captures well the underlying assumptions of realism in general²¹ and classical deterrence theory in particular. When analyzed as a *non-cooperative game* in which binding agreements are not permitted, it mirrors the anarchy condition; as a *non-zero-sum game*, it captures the general understanding among classical deterrence theorists that, in the nuclear age at least, states have a common interest in avoiding war;²² and as a *two-person game*, it starkly reflects the bipolar post-war international system.

¹⁹ We use a simple convention to distinguish between outcomes in a game model and real-world events with the same name: game outcomes are italicized; analogous real-world events are not.

²⁰ For a definition of strategic-form (or normal-form) games, see chapter 2. An informal treatment allows us to describe a wider range of decision-theoretic deterrence theory. As we move on, however, our analysis will become progressively more formal.

²¹ For the general compatibility of game theory and realism, see Jervis (1988a).

²² This consensus took some time to develop and was due, in no small part, to the work of Thomas Schelling. Reflecting perhaps the intensity of the McCarthy period in the United States of the 1950s, almost all of the early applications of game theory analyzed interstate conflicts as zero-sum games. (See, for instance, McDonald and

As well, the players are presumed to be undifferentiated (i.e., “billiard balls”), rational, and egotistical: each most prefers to gain an advantage. Most significant, though, from the point of view of theory construction, is Chicken’s defining characteristic: conflict is the *worst* possible outcome for both sides.

Consider the theoretical implications of this critical, yet not clearly justified *assumption*: by accepting the Chicken analogy, decision-theoretic deterrence theorists perforce presume the “irrationality” of outright conflict;²³ they structure a model in which no “rational” leader would *ever* purposefully choose to resist aggression; and thereby, they presuppose that only irrational and accidental wars are possible. In other words, by uncritically embracing the Chicken analogy, this group of classical deterrence theorists takes as given many of the major *propositions* of structural deterrence theory!

At this juncture one might conclude that decision-theoretic deterrence theorists presume too much, that the assumptions embedded within a Chicken model are so prejudicial that an unimpassioned inquiry into the dynamics of interstate conflict is no longer possible. But such a judgment would be premature. Assumptions, even heroic assumptions, are simply useful devices for facilitating the construction and refinement of theories (Friedman, 1953). Rather, it is the collection of propositions that flows from such theoretical primitives that is telling. In the end, it is the empirical accuracy of these propositions, not the assumptions that led to them, by which a theory must be judged.

Viewed in this light, decision-theoretic deterrence theory serves a particularly useful purpose. By presupposing the world envisioned by structural deterrence theory, the models developed by decision-theoretic deterrence theorists help to specify the logical implications of structural deterrence theory. In other words, these models map out what optimal strategic behavior would be in the world implied by structural deterrence theory. Thus, an evaluation of the theoretical consequences and the empirical accuracy of the models of decision-

Tukey, 1949; McDonald, 1950; Haywood, 1954; Williams, 1954; and Morgenstern 1959, 1961a.) When *The Strategy of Conflict* was re-published in 1980 by Harvard University Press, Schelling remarked in a new preface that the idea that conflict and common interest were not mutually exclusive, so obvious to him, was among the work’s most important contributions.

²³ For instance, a mutually worst outcome cannot be an equilibrium in any sense in any game with strict preference rankings over outcomes.

Theoretical underpinnings

theoretic deterrence theorists can help put structural deterrence theory itself to a more refined test. It is with this noble purpose in mind, then, that we turn to a brief description of the conceptual model lurking beneath the decision-making strand of classical deterrence theory.

Perhaps the easiest way to describe the underpinnings of decision-theoretic deterrence theory is to consider Chicken in light of an expected utility model of blackmail developed by Daniel Ellsberg (1959). Ellsberg's *critical risk* model fully reflects Kissinger's (1994: 481) view that the "art of policy is to create a calculation of the risks and rewards that affect the adversary's calculations." Underlying this model is a set of assumptions common to many decision-theoretic deterrence theorists (Wagner, 1992a).

First is *strategic uncertainty*. Neither player knows for sure which strategy the other will choose. Without a doubt, this lack of information inordinately confounds the decision problem for the players in Chicken. For instance, say that State A knows for sure that State B plans to cooperate. Then its best choice is to defect, because defection yields A's most-preferred outcome and cooperation its second most-preferred outcome. Conversely, if A knows for sure that B plans to defect, then its best choice is to cooperate. (Defection gives A its worst outcome, cooperation its next-worst.) But without certain knowledge of B's choice, State A's optimal choice is unclear.

Second is the *subjectivity* assumption. Although the players are uncertain about each other's behavior, each makes a subjective estimate (based perhaps on intelligence reports, past experiences, prophecies, astrological readings or, as Hans Morgenthau once suggested, hunches) of the other's behavior, expressed as a subjective probability attached to each of the opponent's possible actions. Each player also assesses, again subjectively, the other's (cardinal) utilities for the possible outcomes, i.e., the worth of each outcome to the opponent.²⁴ Of course, a player knows its own utilities.

Finally, there is the *rationality* assumption. Using the probability estimates and the utility assessments, the players act *as if* they were expected utility maximizers. In other words, each player chooses its strategy so as to achieve the highest expected utility. For example, in

²⁴ By utility, we mean von Neumann-Morgenstern utility. (For a discussion, see Morrow, 1994a: ch. 2.) What is important is to be able to interpret the expected utility of a lottery over outcomes as the utility of the lottery itself. For examples, see footnote 25.

Chicken, a player cooperates when the expected utility of cooperation, $E(C)$, exceeds the expected utility of defection, $E(D)$, and defects otherwise.²⁵

In Ellsberg's model, which is based on Chicken, each player's *critical risk* occurs when the expected utilities of its two strategies are equal, that is, when $E(C) = E(D)$.²⁶ This risk is critical in that it represents the maximum risk of conflict (DD) a defecting player is willing to tolerate. At any higher risk level, a rational player simply cooperates. Hence, the lower a player's critical risk, the more likely it is to cooperate; the higher a player's critical risk, the more likely it is to defect.

The calculations seem simple enough, but the choices facing decision-makers involved in a crisis clearly are not. Lurking in the background are two considerable dangers. The most obvious is the very real chance of disaster: if both players stand firm and defect, an "accidental" war (that no one really wants) results. But there is also the risk of losing the advantage by making unnecessary concessions. The rub, of course, is that to avoid one calamity, one must face the other. It is small wonder, then, that during the Cold War period, crises came to be seen as "competitions in risk taking." Everything else being equal, the player with the highest critical risk would "win" the contest.²⁷

Given the risks, what is the best way to play this most dangerous game? Like latter day Machiavellis, decision-theoretic deterrence theorists were at the ready to provide policy-makers with answers to this question, proffering sage advice for managing acute interstate crises. The tactics they suggested were both novel and counter-intuitive. The prescriptions soon gained wide currency in both official

²⁵ To illustrate one way to make this calculation, assume that the ordinal ranks in figure 1.2 represent utilities and that State A believes that there is a 40 percent probability that State B will defect. Then

$$\begin{aligned} E(C) &= 3 (.6) + 2 (.4) = 2.6 \\ E(D) &= 4 (.6) + 1 (.4) = 2.8 \end{aligned}$$

Since $E(D) > E(C)$, State A should defect.

²⁶ In the previous example, A's critical risk is .5 because that is the "crossover" point where $E(C) = E(D)$. Any estimate of the probability that B plans to defect greater than .5 makes it rational for A to cooperate, and conversely. (This calculation follows a formula given by Jervis, 1972. For an alternative method of calculating a player's critical risk, see Snyder, 1972.)

²⁷ Powell's (1990) model reaches the opposite conclusion. For a discussion, see chapter 2.

Theoretical underpinnings

(Kaplan, 1983) and academic circles in the United States, “even though there was little evidence for the validity of the propositions” and even though several recommended tactics were “contrary to common sense” (Jervis, 1979: 289, 292) or appeared “bizarre” (Rapoport, 1992). Indeed, despite empirical, logical, and ethical challenges (Rapoport, 1964; Green, 1966; Young, 1968; George and Smoke, 1974; Smoke, 1977; Snyder and Diesing, 1977; and Zagare, 1987, 1990a), the collected wisdom of decision-theoretic deterrence theory became the conventional wisdom, and remains so to this day.

It is important to keep in mind that all of the strategic imperatives discovered by these “Neo-Clausewitzians,” as Rapoport (1968) pejoratively referred to classical deterrence theorists, flow from the confluence of Ellsberg’s critical risk model and the structural dynamics of Chicken. Jervis (1979: 301), for example, once confessed that he gained insight into deterrence “only by deducing the bargaining tactics that should be effective in the game of Chicken.”

To manage crises successfully, decision-makers were counseled to take actions that decreased the other player’s critical risk, making it more likely that the other player would back down. Snyder (1972) provides an extensive list of manipulative bargaining tactics and a useful scheme for categorizing them. *Critical risk tactics* include both *accommodative moves* designed to make cooperation more attractive for an opponent, and *coercive moves* that make defection more costly and, hence, less attractive. Much more provocative, however, are *commitment tactics* that aim to alter an opponent’s estimate of the probability that one intends to defect.²⁸ The appeal of committing, even probabilistically, to a hard-line strategy in Chicken is obvious: *ceteris paribus*, the higher an opponent’s estimate of the probability that one intends to defect, the lower the opponent’s critical risk and the higher the opponent’s probability of concession – and conversely.

The best way to assure victory during a crisis, then, is to make the opponent believe that concession is impossible. Of course, the same incentive exists for the opponent so that movement from theory to practice is not altogether straightforward. Commitment is an art, not a science (Schelling, 1966: ch. 2). Nonetheless, in coercive bargaining situations, statesmen are counseled to “relinquish the initiative” by

²⁸ See Dixit and Nalebuff (1991: ch. 6) for an informal discussion and a more contemporary listing of commitment tactics.